

A method for detecting the profile of an author

Autores

Jesus Silva, Silvia García, María Alejandra Binda, Fredy Marin Gonzalez, Rosio Barrios, Bellanit Leon Castro, Ligia Castro

Abstract

This paper presents a method for detecting an author's profile using the following two elements: gender and age. This is based on a set of dialogues, written in two languages: English and Spanish, provided for Author Profiling competence within the evaluation forum "Uncovering Plagiarism, Authorship, and Social Software Misuse" (PAN2018). Counts of lexical, semantic, and syntactic characteristics are used to generate a two-phase classification system, which first classifies gender and then age. The results obtained show that, with the amount of data available, it is possible to characterize both the age and gender of an author with an accuracy greater than 50%. However, these values could be improved by having more evidence of information in the training data.

Palabras clave

Supervised Classification, PAN 2018, Gender, Age, Random forest